Amendments to the Specification

Please replace paragraph [0003] with the following amended paragraph.

[0003] The present invention relates to analyzing gene expression. More particularly, the present invention relates to systems and methods for analyzing genes and their expression profiles based upon array analysis.

Please replace paragraph [0034] with the following amended paragraph.

[0034] In an embodiment, the plurality of measured signals correspond to hybridization data used to measure the expression of a gene or a plurality of genes. For example, in an embodiment, the plurality of measured signals comprise a plurality of known DNA sequences hybridized to mRNA isolated from the at least one cell type. In an embodiment, the plurality of known DNA sequences are arranged to form a solid-state array. For example, microarrays such as those commercially available from AFFYMETRIX® ~~Affymetrix~~ (Santa Clara, CA) may be used.

Please replace paragraph [0039] with the following amended paragraph.

[0039] In an embodiment, the ICA algorithm is designed to reduce or minimize computational memory required for matrix analysis. For example, the analysis may comprise hierarchical ICA such that the complexity of the computational analysis is reduced as the analysis proceeds by removing data inputs that have been described at earlier stages of the analysis from the set of data points still remaining to be characterized. Also, several commercially available ICA algorithms known to provide efficient analysis, with significantly reduced memory demands may be used. In an embodiment, the ICA algorithm is FastICA, which runs on ~~MATLAB~~ MATLAB® software (The MathWorks, Inc., Natick MA) may be used. In an embodiment, efficient ICA algorithms may not be as accurate as standard ICA algorithms. Thus, an assessment of the ability of the algorithm to provide the analysis required may be performed using test data prior to utilizing a particular algorithm for the analysis or part of the analysis.

Please replace paragraph [0041] with the following amended paragraph.

[0041] In an embodiment, the present invention further allows for determining the interaction of genes within a gene group. In this way, the interrelationship of individual genes within a particular group of genes is described. In an embodiment, the cross-correlation between genes within a group is determined, wherein a positive cross-correlation comprises the situation in which the expression of one gene in the group is statistically correlated with the expression of a second gene in the same group. In an embodiment, the expression of one of the genes is dependent upon expression of the other gene. For example, the analysis may describe which genes are activated in response to a particular environmental stimulus or pharmaceutical agent. Additionally, the analysis may describe how the expression of one gene in the group affects, or is effected by, the expression of a second gene in the group.

Please replace paragraph [0042] with the following amended paragraph.

[0042] In an embodiment, the relationship between genes within a group is expressed as a mathematical model describing relative levels of gene expression for at least two of the genes in the group. It is understood that gene expression does not occur in isolation and thus, may be influenced by, or reflect the contribution of, a variety of extraneous factors. For example, in an embodiment, the mathematical model that describes expression of genes $y_1, y_2 \ldots y_N$, in a group may include the contribution of at least one environmental factor. Or, the model may include the contribution of time. The model may also include the contribution of noise. For example, in an embodiment, the model comprises the expression $y_i = f(y_1, y_2, \ldots y_N, u_1, \ldots . u_M) + e$, where $y_1, y_2 \ldots y_N$, is the expression of genes 1, 2, . . . N; $u_1, u_2 \ldots u_M$, corresponds to environmental factors 1, 2 . . . M, and experimental noise is defined by e.

Please replace paragraph [0075] with the following amended paragraph.

[0075] DNA microarrays are essentially solid-state grids containing short sequences of DNA of known sequence fixed at a particular position (or address) on the grid (Bassett, D.E., et al., *Nature Genetics*, **21**:51-55, 1999; Hughes T.R., et al., *Curr. Opin. Chem. Biol.*, Feb;**5**:21-5, 2001; Harkin D.P., *Oncologist*, **5**:501-7, 2000; Southern, E., et al.,

Nature Genetics, **21**: 5-9, 1999; Greenberg S.A., *Neurology*, **57**:755-61, 2001; Schulze A., *Nat. Cell. Biol.*, Aug;3(8):E190-5, 2001; Bowtell, D.D., *Nature Genetics*, **21**:25-32, 1999; Devaux, F., et al., *FEBS. Lett.*, **498**:140-4, 2001; Cheung, V.G., et al., *Nature Genetics*, **21**:15-19, 1999; Blohm, D.H., ~~Duggan, D.J.,~~ et al., *Curr. Opin. Biotechnol.*, **12**:41-7, 2001; Hegde P., et al., *Biotechniques,* Sep;29(3):548-50, 2000, Duggan, D.J., et al., *Nature Genetics*, **21**:10-14, 1999, Hacia, J.G., *Nature Genetics*, **21**:42-47, 1999. In some cases, the DNA sequences are short fragments of DNA generated from a library. Alternatively, some arrays comprise oligonucleotide sequences (short fragments of DNA less than 50 nucleotides long) each of which may only differ by one base pair (commercially available from AFFYMETRIX® ~~Affymetrix~~, Santa Clara, CA).

Please replace paragraph [0101] with the following amended paragraph.

[0101] A variety of systems known in the art may be used for image analysis (110) and compiling the data (120). For example, where the mRNA is labeled with a fluorescent tag, an fluorescence imaging system (such as the microarray processor commercially available from AFFYMETRIX® ~~Affymetrix~~, Santa Clara, CA) may be used to capture, and quantify the extent of hybridization at each address. Or, in the case where the mRNA is radioactive, the array may be exposed to X-ray film and a photographic image made. Once the data is collected, it may be compiled to quantify the extent of hybridization at each address as for example, using software to convert the measured signal to a numerical value.

Please replace paragraph [0102] with the following amended paragraph.

[0102] In an embodiment of the present invention, the first step of signal analysis after the imaging data has been collected (110) and collated and/or transformed into a quantitative signal (120), is to filter noise from the data (130). In an embodiment, filtering may be conducted via a thresholding mechanism in which genes whose expression levels are always less than the noise level for the microarray are removed from the pool. Alternatively, variance normalization techniques may be applied to the data. Such variance normalization may include, but is not limited to, statistical techniques on based on the normalization around mean, median, start point, or the

endpoint. Software for performing data normalization as a means to remove noise from data sets includes, but is not limited to, ~~MATLAB~~ MATLAB® (The MathWorks, Natick, MA) and SAS (SAS Institute, Inc., Cary, NC).

Please replace paragraph [0112] with the following amended paragraph.

[0112] There are many types of ICA algorithms that may be used for the analysis techniques of the present invention. For example, linear ICA algorithms may be employed. Linear algorithms may be preferred for the analysis of large data sets of microarray data ~~(Berger, J.A., et al., Microarray Data Using Independent Component Analysis, *Proceedings of the International Symposium on Control, Communications, and Signal Processing*, Hammamet, TUNISIA, March 21-24, 2004)~~. Also, non-linear ICA algorithms (such as P. Pajunen, A. Hyvärinen and J. Karhunen "Non-Linear Blind Source Separation by Self-Organizing Maps" *Proc. Int. Conf. on Neural Information Processing*, Hong Kong, pp. 1207-1210, 1996) may also be employed. Non-linear algorithms may be better suited to smaller data sets comprising fewer gene groups due to the ability of these algorithms to analyze non-linearly combined signals.

Please replace paragraph [0114] with the following amended paragraph.

[0114] Thus, in an embodiment, the ICA algorithm may comprise a means to reduce the computational intensity required for each step. For example, an ICA algorithm called FastICA (available from the Helsinki University of Technology, Laboratory of Computer and Information Science) may be implemented in ~~MATLAB~~ MATLAB® 6.0 (The MathWorks, Inc., Natick, MA) using a 1GHz AMD ATHLON PROCESSOR® ~~Athlon PC~~. Although not as accurate as the original ICA, FastICA provides a much faster technique when large datasets are used.

Please replace paragraph [0141] with the following amended paragraph.

[0141] Embodiments of computer-readable media include, but are not limited to, an electronic, optical, magnetic, or other storage or transmission device capable of providing a processor with computer-readable instructions. Other examples of suitable media include, but are not limited to, a floppy disk, CD-ROM, magnetic disk, memory chip,

ROM, RAM, an ASIC, a configured processor, all optical media, all magnetic tape or other magnetic media, or any other medium from which a computer processor can read instructions. Also, various other forms of computer-readable media may transmit or carry instructions to a computer, including a router, private or public network, or other transmission device or channel, both wired and wireless. The instructions may comprise code from any computer-programming language, including, for example, C, ~~C#, Visual Basic, Visual Foxpro~~, VISUAL C#®, VISUAL BASIC®, VISUAL FOXPRO®, Java, and JavaScript.

Please replace paragraph [0148] with the following amended paragraph.

[0148] Once the data has been collected (i.e., using the imaging system or other type of data collection system), it may be compiled (120) and/or transformed if necessary using any standard spreadsheet software such as Microsoft Excel, FOXPRO® ~~FoxPro~~, Lotus, or the like. In an embodiment, the data are entered into the system for each experiment. Alternatively, data from previous runs are stored in the computer memory (160) and used as required.

Please replace paragraph [0151] with the following amended paragraph.

[0151] In the next step, iterative ICA (140) is performed as described above. In some cases, the user may want to input variables or constraints for the analysis, as for example, where the number of gene groups is known. Also, the user may choose which type of ICA software may be employed for the analysis. Examples of commercially available processing software suitable for microarray analysis include, but are not limited to ~~MATLAB~~ MATLAB® (The MathWorks, Natick, MA), and SAS (SAS Institute, Inc., Cary, NC). For example, in one embodiment of the present invention, a prototype was coded using ~~MATLAB~~ MATLAB® (without using the MATLAB® ~~MATLAB's~~ bioinformatics toolbox or any other ~~MATLAB~~ MATLAB® function specialized for microarray processing). Thus, as described above, in some situations, FastICA may be the software of choice. Other situations, however, may require a more rigorous ICA program to be used.

Please replace paragraph [0153] with the following amended paragraph.

[0153] FastICA (available from Helsinki University of Technology) implemented in ~~MATLAB~~ MATLAB® 6.0 (The Mathworks, Inc., Natick, MA) was used for the analysis of gene expression data relating to bone healing. Although not as accurate as the traditional ICA methods, FastICA provides a much faster technique when large datasets are used. The hardware was a 1GHz AMD ~~Athlon PC~~ ATHLON PROCESSER® with 768 Mb memory. The input data was a series of four microarray measurements for the expression levels of genes of young rats during the bone fracture healing process. The data also included a starting non-fracture measurement.

Please replace paragraph [0154] with the following amended paragraph.

[0154] The largest data set included 8,799 data points. For the removal of noise, the data from the control (the non-fracture reading) was subtracted from each subsequent data point (after fracture and during the healing process). Using a simple (i.e., non-optimized) filtering technique, series with changes less than 100 linear units of expression level were discarded. In the matrix array analyzer used ~~(Affymetrix)~~ (AFFYMETRIX®), expression levels below 100 are believed to be caused by machine noise and other sources of noise. Subtraction of noise left 4,315 series to consider. Due to the use of auto-correlation techniques for filtering (as discussed above) the filtering process is optimal and eliminates the noise-like patterns that have no significant biological basis.

Please replace paragraph [0158] with the following amended paragraph.

[0158] Figure 3 shows a composite of all 94 signals for these genes. Although the figure is too compressed in some regions to show correlations for many of the genes, some of the outer sequences can be seen to have highly correlated (or anti-correlated) expression profiles. The gene names used in Table 1 are the standard names used in public databases such as the data base available as the rat genome database at the Medical College of Wisconsin website ~~at http://rgd.mcw.edu~~.